20

Language Documentation & Conservation Special Publication No. 15
Reflections on Language Documentation 20 Years after Himmelmann 1998
ed. by Bradley McDonnell, Andrea L. Berez-Kroeker & Gary Holton, pp. 202–209
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/24822

Reflections on linguistic fieldwork

Claire Bowern
Yale University

In this reflections piece, I draw upon my experience as a fieldworker in Australia, a linguist who also works with archival materials spanning 150 years, and a linguist whose work includes both documentary and descriptive aspects. I center this piece around three questions about aspects of fieldwork that have changed since the publication of Himmelmann (1998). The first is what we collect — that is, have our field methods changed? The second question concerns the documentation we produce — is it different? Thirdly, are there features of Himmelmann's manifesto which were the products of its time, and has academia changed? Arguably in all cases that there has been change for the better, but we still have some way to go, and that some of the original formulation of a dichotomy between documentation and description are counterproductive.

1. Introduction[1] In this reflections piece, I draw upon my experience as a fieldworker in Australia, a linguist who also works with archival materials spanning 150 years, and a linguist whose work includes both documentary and descriptive aspects. I began fieldwork in 1999, and so my entire professional career as a fieldworker has been in what we might call the "post-Himmelmann" era of language documentation. That is, it has been conducted in the intellectual environment of explicit discussions of field methodology, of documentation as a practice distinction from linguistic analysis, and with in-depth discussion of what it means to work ethically with communities, speakers, and language data.

I center this piece around three questions about aspects of fieldwork that have changed since the publication of Himmelmann (1998). The first is what we collect — that is, have our field methods changed? I argue that field methods have, indeed, changed in several ways. There is more interdisciplinary work; more collaboration with language communities, and more recognition of what needs to go into a documentation project for communities. However, there is still a lot more to say about what we describe when linguists work "on a language". Himmelmann (1998:161-63, 166) defines the basic object

[1]Many thanks to two anonymous reviewers whose comments helped me refine discussion of several topics presented here.

ISBN: 978-0-9973295-3-7

of description as being "observable *linguistic behavior*, manifest in everyday interaction between members of the speech community, and 2) native speakers' *metalinguistic knowledge*, manifest in their ability to provide interpretations and systematizations for linguistic units and events." I argue that this does not fit all cases and we should not confine ourselves to field situations where both these objects of description are accessible.

The second question concerns the documentation we produce — is it different? Do we see more corpora (primary descriptive materials) being produced and published? Have we changed what we are doing with the primary resources? I argue that much greater availability of digital recording devices has been the driver of change here, along with a greater focus on archiving as distinct from publishing, but changes are slow.

Thirdly, are there features of Himmelmann's manifesto which were the products of its time, and has academia changed? Again, there has been change: the dichotomy between 'community' and 'linguist' is not quite as pronounced as it was 20 years ago; there are more linguists from endangered language communities working on their own languages, and the partnership that Wilkins (1992) discusses for Australia is a more usual way to conduct research, in both the US and Australia and other parts of the world. There is less work that takes no account of community dynamics and pressures. But we could, and should, be doing better. These points have relevance to the objects of documentation, as well as what is done with documentary materials.

2. Have field methods changed? I see several important changes in the type of fieldwork typically undertaken by linguists over the past twenty years. The first is the increasing use of semi-structured elicitation tasks in basic language documentation. Perhaps most famous are the Max Planck Institute (Nijmegen) field manual kits released over the period 1992-2010.[2] The kits provide visual stimuli, questionnaires, and structured, consistent tasks for use with speakers. Although originally developed for MPI-internal comparative/typological research projects, both the specific stimuli and the general approach have been used as a way of getting controlled data without the prompting of translation-based sentences or the need for speakers to be fluent in a contact language. Further discussion of these methods can be found in Bochnak & Matthewson (2015) and Cover & Tonhauser (2015) amongst others.

The second difference has been the emphasis on the collection of conversational and natural data as part of a documentation project, even when the main goal of the language documentation project is not conversation analysis. While linguists have long made use of a variety of methods for gaining information about the language (not least, participant observation and learning the language; cf. Hale (2001) "do whatever it takes to learn the language"), linguists are now both more explicit about their documentation methods and are using more approaches consistently and deliberately. Indeed, so concerned is Thieberger (2012) about the type of material that is missing from traditional guides that only one chapter in that handbook (Mosel 2012's 'guide to the guides') covers what might be called the core of "traditional" fieldwork. While Abbi (2001), Bowern (2008), Crowley (2007), and others are all different books, they do cover much of the same general material, with a focus on language documentation through subfields of linguistics.

Impressionistically, there has been a methodological "smoothing" over the last twenty years. As a graduate student, I remember heated debates about whether descriptions

[2]These handbooks were originally distributed in print with CDs. They are now available from `http://fieldmanuals.mpi.nl`.

should be based on elicited data or conversational data (alone); each camp firmly convinced of the uselessness of the type of data produced by the other's methods. In brief, elicited data was considered too contaminated by the meta-language of elicitation to be a "true" reflection of data from the language, while conversational data was considered too unstructured and incomplete to be useful at elucidating the internal grammatical competence of an individual. We have a better and more nuanced understanding of the advantages and disadvantages of different types of data — that elicitation and translation provides better evidence of the possible and impossible structures of languages, while natural interactions provide a valuable source of structural data, as well, of course, data about the linguistic behavior itself (more on that below). No doubt, this has come about through more outlets for explicit discussion of field methodology, such as through the pages of journals such as *Language Documentation & Conservation* and *Language Documentation and Description,* as well as recent handbooks (Austin & Sallabank 2011; Thieberger 2012; Chelliah & de Reuse 2010).

In Australia, there are more genuinely interdisciplinary projects. Evans (2012) provides an overview, and Thieberger (2012) has numerous chapters about linguistic knowledge of cultural practices ranging across the natural world. An example of one from Australia is Glenn Wightmann's ethnobiological collaborations with communities across the Top End of Australia (for example, 1994; Wightman, Roberts & Williams 1992). It must be noted, though, that interdisciplinary work isn't new in Australia. Some of the earliest intensive academic fieldwork, such as the Cambridge Expedition to the Torres Strait of 1898, was "interdisciplinary" in that they included scholars from many fields who worked together.

Perhaps the biggest difference between field methods from twenty years ago and today is the use of digital recording technology. Being able to record digitally has made a huge difference to the amount of material that can be recorded, to the workflow for processing recordings, and to the type of work that can be done with such material. The limitations of analog media for recording included the expense of the tapes, the amount of space they took up in luggage to and from the field, the difficulty of making backups (usually they had to be backed up in real time, unless one had a tape deck which could play and record at faster than 1x speed), the cumbersomeness of using them for transcription (play a sentence, rewind, play again), and the fragility of the media (don't leave them in the car on a hot day, or near a magnet). Let alone video. As a result, linguists now record a lot more of their sessions, and the audiovisual component of a documentation project is the richer for it. However, extensive audiovisual recordings bring the problems that come with multiple gigabytes of data, the bottleneck of transcription, and material of incidental relevance to the project. But these problems are small compared to the advantages of being able to time-align recordings with transcriptions, being able to search over multiple corpora and sessions within a second, and being able to extract and manipulate data for phonetic analysis.

In summary, we have changed what we collect, but not, I think, because (or purely because) of the language documentation/description divide which Himmelmann (1998) focused on. Rather, tools like the MPI field manuals have brought to greater prominence the role of semi-structured data gathering (Bowern 2008); we have better procedures for interdisciplinary data gathering, and digital data collection and processing tools have reshaped workflows. (For a view that ties this shift to Himmelmann more closely, see Good 2011.) Semi-structured data, however, is seldom useful for maintenance of culture, or linguistically mediated cultural practices.

Before moving to the second question, we should consider the focus of documentation. Himmelmann's focus is on documenting linguistic behavior as the way to document the language. That is, he makes an explicit difference between the recording of linguistic forms which provide evidence for language structures, and the behavior of individuals when talking. I associate this focus particularly with linguists trained within the MPI Nijmegen tradition, and the approach is well illustrated by the field manuals mentioned above. However, this dichotomy is problematic for some field situations. For example, I typically work with speakers who do not use the language we are documenting every day. We have discussions about translations, speakers produce words, sentences, and connected speech in the language, but we document language and linguistic forms with no assumption that what we are doing transfers to a more general set of speech practices. Those speech practices are gone; they've been replaced by a very different linguistic ecology. The majority of the time that the language is used is in these linguistic sessions—that is, the documentation *is* the primary linguistic ecology for these languages these days.

3. Has the production of documentation changed? While our data storage methods have changed greatly with digital recorders, and some types of linguistic data are much easier to collect than they used to be, my impression is that the core of documentation—at least for the linguistic community—is still based around categories of traditional grammar and functional typology. We are getting better data, from more varied sources; we are paying more attention to variation. Handbooks and publications like Thieberger (2012) and Bochnak & Matthewson (2015) have made us more aware of the methods we employ in field work and how to collect data better. But our publications are little changed from the grammars, dictionaries, and texts of 50 years ago. Even the digital grammar collections, like Pacific Linguistics' online *Asia Pacific Linguistics* documentation series, are essentially print books online. We could be doing a lot more to take advantage of the possibilities that digital media allows, such as audiovisual and text linking, or non-linear presentation.

Moreover, since Wilkins' (1992) article on linguistic research under Aboriginal control, a series of papers on ethics, community involvement, and language pedagogy have made documentary linguists more aware of issues of audience, of the difference between pedagogical materials and descriptive grammars, of the ways in which linguistic terminology produces barriers to understanding, and how we can partner with education specialists to improve the materials we produce for and with communities (cf. Czaykowska-Higgins 2009). Simply put, a "community contribution" in the context of an endangered language documentation project is no longer satisfied by a copy of a $300 reference grammar in the local library. While we are still very focused on "the community" (without appropriate recognition that communities are groups of individuals, not all of whom may agree with one another), we are doing a better job at recognizing what an appropriate and meaningful contribution to a community might look like, and that such contributions will differ depending on the field site (see further Dobrin & Schwartz 2016). Now that the first generation of linguists trained under Wilkins' (1992) model are now senior members of the field, we can see the ways in which (for Australia at least) we have lasting recognition that community linguists and language activists are the crucial drivers of language projects.

Another big difference from 20 years ago is the greater emphasis on archiving, the ethics and responsibilities that linguists have to archive their data; the distinction between archiving and publication (even publication on the web, which was still rare in 1998).

Australia was ahead of the game here, so we see less difference over the last twenty years than in other places. Australia already had an excellent (print) archive, the library of the Australian Institute for Aboriginal and Torres Strait Islander Studies (AIATSIS). Other historical collections are held in regional Museums and the National Library of Australia, but the AIATSIS archive is unique in having an extensive and continuing collection of both print and audiovisual resources for Australian languages. Australia also had ASEDA (Aboriginal Studies Electronic Data Archive), which unfortunately no longer exists (much of its holdings has been transferred to AILEC, housed by AIATSIS). Unfortunately, Australia's digital archiving is perhaps now somewhat behind, compared to AILLA and the Berkeley Survey of California and other Indian Languages. Recent digital field archives for Australian documentation projects are spread between ELAR, Paradisec, and DoBeS (sometimes in several of these archives simultaneously). I suspect a great deal of digital documentation material (particularly secondary analytical materials such as conference handouts, slides, or posters, and data from language revitalization events and language classes) is not being archived.[3]

4. Has academia changed? Himmelmann (1998) places the linguist at the center of the documentation process. One difference from 20 years ago is the blurring of lines between data and analyses generated by the linguist and those produced by native speakers. A great deal more language documentation is conducted either by linguists who are also native speakers, or with non-native speaker linguists in conjunction with speakers. However, as Hill (2002) and Davis (2017) have shown, linguists talking about endangered languages still do so in a way that constructs barriers to community members. As linguists continue to discuss ways in which our field could improve, a greater commitment to diversity and inclusion (including the ways in which linguists' practices end up excluding the very people they are aiming to hire) should feature.

Students get much better training in field methods, including archiving, ethics, data, etc. When I started fieldwork, the only textbooks were from the 1960s, apart from Vaux & Cooper (1999), which was problematic for work that was community-based. We now have a wealth of material about what fieldwork is, how to do it, what the ethical implications are, and how to deal with data throughout the documentation process.

5. Further reflections and conclusions A few other points are warranted. Perhaps most important is Himmelmann's definition of language documentation (Himmelmann 1998:166). As briefly discussed above, Himmelmann distinguishes between the recording of linguistic behavior and the recording of linguistic judgments. This is problematic. First, it overlooks the fact that for undocumented or underdocumented languages where the linguist has limited field time, it may be most expedient to structure the language documentation around the analytical results (or the description of another, closely related language) or around the compilation of a dictionary. In fact, several fieldwork books advocate working this way, at the same time as endorsing the approach of a separation of language documentation and description. A wordlist with example sentences is a good way to get enough preliminary data for a sketch (compare Hale 2001). I suspect that linguists, on the whole, plan their documentary activities—at least initially—around the end products. Moreover, data collected without an aim (or hypothesis) can be at

[3]https://zenodo.org/communities/australianlanguages/ is a free community portal for Australian languages where material of this type can be uploaded, particularly for the archiving of "grey" literature.

best problematic, or at worst, useless. It might be fine for some purposes, but specific hypotheses usually require specific types of data. These goals should be consistent with community expectations of the documentation project as well. For example, if a linguist is brought in under the assumption that the project will document language in use, they should not simply make a wordlist.

"As long as collection and analysis are considered part of a single, uniform, project, the collection activity is likely to be (relatively) neglected" (Himmelmann 1998:164). More primary data are available, thanks to online archives. However, it needs to be acknowledged that primary data are still difficult to use without familiarity with the language (or one closely related). For example, as beautifully laid out and user-friendly as the online Ainu[4] corpus is, realistically, for most research purposes, I will be looking for an Ainu grammar. And because there are so many languages and so few linguists working on them, it's often the case that the only person with the requisite knowledge to use raw data from a corpus collection is the linguist who collected it in the first place. So, linguists end up using the secondary sources anyway, even if the raw data are available. Another example is the Chirila database of Australian wordlists (Bowern 2016); because the scope of the material is Australia-wide, we could not realistically work from audio or unprocessed field notes, but have had to initially prioritize analyzed or at least partially processed sources. There are, however, other use cases where availability of raw materials is preferable. Communities using and adapting materials, for example, are likely to need the primary data.

How much of the changes discussed here are due to the influence of Himmelmann (1998) alone? It's hard to say. Certainly, the paper has been influential, highly cited, much discussed in the literature on language documentation, and has been accompanied by other very influential publications on fieldwork and recording language. Yet it came at the right time for other changes in academia too — particularly the influence of digital resources on documentary methods. Benchmark papers like this let us see how far we've come, but they should not either prevent us from seeing what came before, or stop us from re-imagining 21st Century fieldwork as documentation and description that works with and enhances communities even further.

[4]http://ainucorpus.ninjal.ac.jp/corpus/en/

References

Abbi, Anvita. 2001. *A manual of linguistic field work and structures of Indian languages*. Munich: Lincom Europa.

Austin, Peter K. & Julia Sallabank (eds.). 2011. *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press.

Bochnak, Ryan & Lisa Matthewson. 2015. *Methodologies in Semantic Fieldwork*. Oxford University Press.

Bowern, Claire. 2008. *Linguistic fieldwork: A practical guide*. Basingstoke: Palgrave Macmillan.

Bowern, Claire. 2016. Chirila: Contemporary and Historical Resources for the Indigenous Languages of Australia. *Language Documentation & Conservation* 10. 1-44.

Chelliah, Shobhana & Willem de Reuse. 2010. *Handbook of Descriptive Linguistic Fieldwork*. New York: Springer.

Cover, Rebecca & Judith Tonhauser. 2015. Theories of meaning in the field: Temporal and aspectual reference. In Ryan Bochnak & Lisa Matthewson (eds.), *Methodologies in Semantic Fieldwork*, 306–349. Oxford University Press.

Crowley, Terry. 2007. *Field linguistics: A beginner's guide*, ed. by Nick Thieberger. Oxford University Press.

Czaykowska-Higgins, Ewa. 2009. Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian indigenous communities. *Language Documentation & Conservation* 3(1). 15-50. `http://hdl.handle.net/10125/4423`

Davis, Jenny L. 2017. Resisting rhetorics of language endangerment: Reclamation through Indigenous language survivance. *Language Documentation and Description* 14. 37-58.

Good, Jeff. 2011. Data and language documentation. Cambridge University Press. In Peter Austin & Julia Sallabank (eds.), *The Cambridge handbook of endangered languages*, 212–234. Cambridge: Cambridge University Press.

Hale, Kenneth L. 2001. Ulwa (Southern Sumu): The beginnings of a language research project. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*, 76–101. Cambridge: Cambridge University Press.

Hill, Jane H. 2002. "Expert Rhetorics" in Advocacy for Endangered Languages: Who Is Listening, and What Do They Hear? *Journal of Linguistic Anthropology* 12(2). 119–133.

Himmelmann, Nikolaus. 1998. Documentary and descriptive linguistics. *Linguistics* 36(1). 161–196.

Thieberger, Nicholas (ed.). 2012. *The Oxford handbook of linguistic fieldwork*. Oxford University Press.

Vaux, Bert & Justin Cooper. 1999. *Introduction to linguistic field methods*. Munich: Lincom Europa.

Wightman, Glenn M. 1994. *Gurindji ethnobotany: Aboriginal plant Use from Daguragu, Northern Australia*. Darwin: Conservation Commission of the Northern Territory.

Wightman, G., J. G. Roberts, & L. Williams. 1992. Mangarrayi ethnobotany aboriginal plant use from the Elsey area Northern Australia. *Northern Territories Botanical Bulletin* 15.

Wilkins, David. 1992. Linguistic research under Aboriginal control: A personal account of fieldwork in Central Australia. *Australian Journal of Linguistics* 12(1). 171–200.

Claire Bowern
claire.bowern@yale.edu
orcid.org/0000-0002-9512-4393